Protist, Vol. 155, 407–422, December 2004
http://www.elsevier.de/protist
Published online 3 December 2004

Protist

ORIGINAL PAPER

Early Evolution within Kinetoplastids (Euglenozoa), and the Late Emergence of Trypanosomatids

Alastair G. B. Simpson[a,1], Erin E. Gill[b], Heather A. Callahan[c], R. Wayne Litaker[d], and Andrew J. Roger[b]

[a] Canadian Institute for Advanced Research, and Department of Biology, Dalhousie University, Halifax, B3H 4J1, Canada
[b] Canadian Institute for Advanced Research, and Department of Biochemistry, Dalhousie University, Halifax, B3H 1X5, Canada
[c] United States Department of Agriculture-Agricultural Research Service, George E. Brown, Jr. Salinity Laboratory, 450 West Big Springs Road, Riverside, California 92507, USA
[d] Center for Coastal Fisheries and Habitat Research, National Ocean Service, National Oceanic and Atmospheric Administration, Beaufort, North Carolina 28516-9722, USA

Submitted April 30, 2004; Accepted June 20, 2004
Monitoring Editor: Larry Simpson

Many important relationships amongst kinetoplastids, including the position of trypanosomatids, remain uncertain, with limited taxon sampling of markers other than small subunit ribosomal RNA (SSUrRNA). We report gene sequences for cytosolic heat shock proteins 90 and/or 70 (HSP90, HSP70) from the potentially early-diverging kinetoplastids *Ichthyobodo necator* and *Rhynchobodo* sp., and from bodonid clades '2' (Parabodonidae) and '3' (Eubodonidae). Some of the new cytosolic HSP70 sequences represent a distinct paralog family (HSP70-B), which is related to yet another paralog known from trypanosomatids (HSP70-C). The (HSP70-B, HSP70-C) clade seemingly diverged before the separation between kinetoplastids and diplonemids. Protein phylogenies support the basal placement of *Ichthyobodo* within kinetoplastids. Unexpectedly, *Rhynchobodo* usually forms the next most basal group, separated from the clade '1' bodonids with which it has been allied. Bootstrap support is often weak, but the possibility that *Rhynchobodo* represents a separate early-diverging lineage within core kinetoplastids deserves further testing. Trypanosomatids always fall remote from the root of kinetoplastids, forming a specific relationship with bodonid clades 2 (and 3), generally with strong bootstrap support. These protein trees with improved taxon sampling provide the best evidence to date for a 'late' emergence of trypanosomatids, contradicting recent SSUrRNA-based proposals for a relatively early divergence of this group.

Introduction

The kinetoplastids are a particularly well-known group of heterotrophic protists. The trypanosomatids are one of the most successful groups of specialist parasites on earth, and include several organisms of tremendous medical and economic importance (e.g. *Trypanosoma* spp. – causing Chagas' disease and sleeping sickness in humans; *Leishmania* spp. – causing kala azar and other leishmaniases). Thanks to their bizarre cytology, molecular biology and genome organisation (Donelson et al.

[1] Corresponding author;
fax 1 902 423 3736
e-mail alastair.simpson@dal.ca

1434-4610/04/155/04-407 $ 30.00/0

1999) trypanosomatids are also the objects of considerable basic scientific interest. Kinetoplastids other than trypanosomatids are referred to as 'bodonids'. Some of these are also parasitic, with some *Ichthyobodo* and true trypanoplasms (e.g. *Cryptobia salmositica*) causing significant disease in commercial fish (Woo and Poynton 1995). However, the bulk of bodonid diversity is made up of free-living phagotrophs, which as a group can be dominant bacterivores in benthic ecosystems (Arndt et al. 2000).

A clear picture of the phylogenetic relationships amongst kinetoplastids is critical to understanding the peculiar biology of trypanosomatids (and other important taxa) from an evolutionary perspective (Lukeš et al. 2002; Maslov et al. 1994; Simpson et al. 2000). Surprisingly, the deep-level evolutionary history of the group has remained poorly resolved, despite substantial efforts in phylogenetic analysis, predominantly using nuclear small subunit ribosomal RNA (SSUrRNA) gene sequences (e.g. Atkins et al. 2000; Doležel et al. 2000; Lukeš et al. 1997; Wright et al. 1999). SSUrRNA studies with ever-improving taxon representation and analysis methods have helped answer some phylogenetic questions, for example, the major relationships amongst trypanosomatids (Hollar et al. 1998; Lukeš et al. 1997; Merzlyak et al. 2001; Stevens et al. 1999) and the polyphyly of the major 'bodonid' taxa *Bodo* and *Cryptobia* (Callahan et al. 2002; Doležel et al. 2000; Hughes and Piontkivska 2003a; Moreira et al. 2004). Nonetheless, most deep nodes within kinetoplastids remain uncertain, principally because it has been extremely difficult to reliably root the kinetoplastid subtree with this marker. In SSUrRNA trees the branch joining kinetoplastids to their closest relatives is extremely long, while most deep internal branches within kinetoplastids are relatively short (Simpson et al. 2002). This suggests that most of the historical signal for the deepest divergences within kinetoplastids has been erased by substitutional saturation, and that long branch attraction to this basal branch, or other artefact, may be responsible for whatever phylogenetic patterns are observed (Moreira et al. 2004; Simpson et al. 2002). Certainly, published analyses place the root of kinetoplastids with or within almost every major cluster identified in unrooted trees, including trypanosomatids (Doležel et al. 2000; Hughes and Piontkivska 2003a; Lukeš et al. 2002; Marin et al. 2003; Simpson et al. 2002).

In recent years there have been two significant improvements in molecular approaches to understanding kinetoplastid evolution. Firstly, increased sampling has finally uncovered two new clades that appear to 'break' the long stem branch for kinetoplastids in SSUrRNA gene trees. The first clade includes the highly reduced *Perkinsiella*-like endosymbionts of paramoebid amoebae, and the fish ectoparasite *Ichthyobodo,* which together form the immediate sister group to all other characterised kinetoplastids (Callahan et al. 2002; Dyková et al. 2003). The second clade is the next closest group and, so far, contains only environmental sequences from deep marine benthos (López-Garcia et al. 2003). However, it is unclear whether the addition of these probable basal groups significantly improves the phylogenetic resolution within the 'apical' or 'core' kinetoplastids (= Metakinetoplastina), as even the most detailed analyses provide weakly supported relationships that are sensitive to outgroup taxon sampling (von der Heyden et al. 2004; Moreira et al. 2004). Interestingly, in some of these analyses (especially those where these two new basal groups are used as the only outgroups), trypanosomatids fall as a clade at the base of Metakinetoplastina, leaving all 'apical bodonids' (= 'core bodonids') as a single clade (von der Heyden et al. 2004; López-Garcia et al. 2003; Moreira et al. 2004).

The second improvement has been the emergence of nuclear-encoded protein data sets with useful taxon sampling within bodonid kinetoplastids. There are now two such markers – the cytosolic isoforms of heat shock proteins 70 and 90 (HSP90 and HSP70) (Simpson and Roger 2004; Simpson et al. 2002). In contrast to SSUrRNA analyses, the branch joining kinetoplastids to other taxa is of moderate length with these markers. Trees estimated from protein sequences always divide studied kinetoplastids into four clades – bodonid clades '1', '2' and '3', and trypanosomatids – and usually place 'clade 1' at the base of the tree relative to clade 2, clade 3 and trypanosomatids (recently, clades '1', '2' and '3' have been named Neobodonidae, Parabodonidae and Eubodonidae respectively – see Moreira et al. 2004). However, the deep branches within HSP90 and HSP70 trees can be weakly supported, and in the case of HSP90, an alternative topology was recovered throughout one series of analyses (Stechmann and Cavalier-Smith 2003). Perhaps most importantly, taxon sampling has been greatly inferior to that available with SSUrRNA. In particular, the deepest identified kinetoplastids in SSUrRNA trees, such as *Ichthyobodo,* have not been included, and neither has *Rhynchobodo,* a taxon that branches with clade 1 kinetoplastids in most SSUrRNA trees (Doležel et al. 2000; von der Heyden et al. 2004; Moreira et al. 2004; Nikolaev et al. 2003; Simpson et al. 2002). This limited taxon sampling weakens the explicatory power of the estimated trees, and could call into question the phylo-

genetic accuracy of the analyses across the larger group of interest.

In this study we aim to reconcile and test the insights from these two disparate molecular data sets by improving the taxon sampling of the existing protein alignments. We report HSP90 sequences for the potentially deep-branching kinetoplastids *Ichthyobodo* and *Rhynchobodo*. We also report an HSP70 sequence from *Rhynchobodo*, and further sequences of this marker from clade 2 and clade 3 bodonids, revealing the existence of deep paralog families within Euglenozoa. Our analyses confirm the basal placement of *Ichthyobodo*, and open the possibility that *Rhynchobodo* might diverge as an independent lineage prior to clade 1. Our study provides the best phylogenetic evidence to date for a 'late' origin of trypanosomatids inside Metakinetoplastina, nested within a paraphyletic radiation of 'core bodonids'.

Results

HSP70 Paralogs

Our alignment of cytosolic HSP70 proteins includes several divergent sequences from kinetoplastids

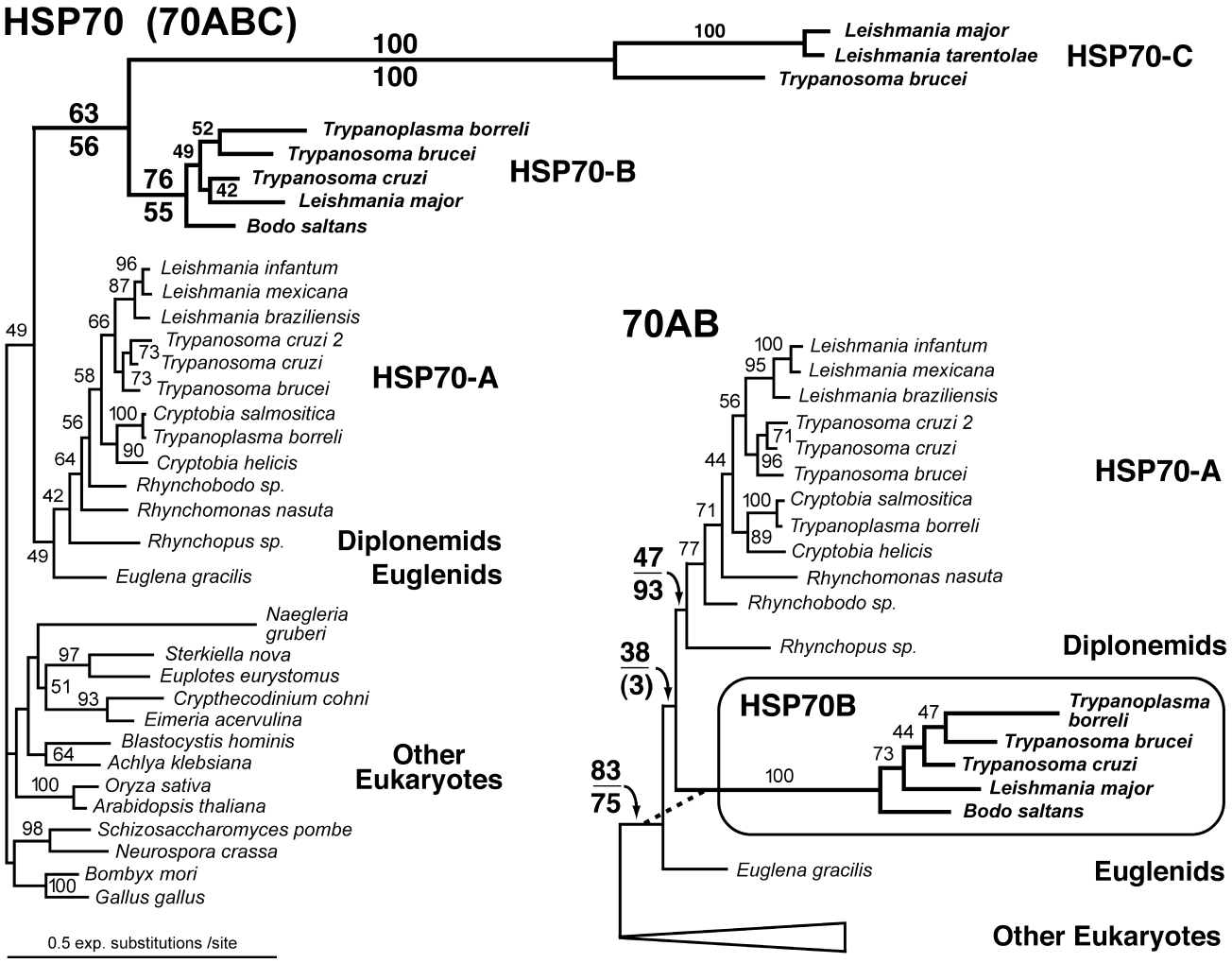

Figure 1. Maximum likelihood trees for HSP70 proteins including the divergent putatively Euglenozoa-specific paralogs HSP70-B and HSP70-C. The tree to the left is estimated from the 70ABC data set, with both HSP70-B and HSP70-C included. The tree to the right is estimated from the 70AB data set, which does not include HSP70-C. The ML topology of the 'other eukaryotes' subtree in the 70AB analysis is the same as that from 70ABC. In both trees, important bootstrap values are shown in large type (ML above, MLdist below – support values for branches not present in the best MLdist tree are given in parentheses). Maximum likelihood bootstrap values for other branches are shown when >40%. The best MLdist tree for 70AB places HSP70-B as the sister to other sequences from Euglenozoa (dashed line). The scale bar (bottom left) applies to both trees.

that appear to represent two 'new' distinct paralogs. The first is represented by *Leishmania major hsp70.4*, which encodes a constitutively expressed protein originally called cp70.4 (Searle and Smith 1993). This gene is also present in the genomes of *Trypanosoma* spp. Hereafter, we refer to the products as 'HSP70-B'. Protein BLAST queries of the Genbank NR database indicate that our new HSP70 sequences from *Bodo saltans* (clade 3) and *Trypanoplasma borreli* (clade 2) are actually HSP70-B. The second paralog includes a constitutively expressed protein encoded by *hsp70.1* in *Trypanosoma brucei* (Lee et al. 1990). Similar genes are also present in *Leishmania* spp. (Brochu et al. 2004). Hereafter we refer to the protein as 'HSP70-C'. Other previously identified ('normal') cytosolic HSP70 sequences from kinetoplastids are referred to as 'HSP70-A'

We report protein-level phylogenetic analyses of two data sets including one or both of the two 'new' paralogs together with HSP70-A and cytosolic HSP70 sequences from other eukaryotes. The first data set, '70ABC' includes both HSP70-B and HSP70-C. The second data set, '70AB' includes HSP70-B only. All analyses reported here model among-site rate variation by a discrete approximation of a gamma distribution, and employ both a maximum likelihood (ML) method, and a least-squares optimisation based on maximum likelihood estimates of pairwise evolutionary distances (MLdist).

Both the ML and MLdist trees for the 70ABC data set place recover HSP70-B and HSP70-C as sister clades (Fig. 1). The extremely divergent HSP70-C sequences form a strongly supported clade. The HSP70-B clade is quite strongly supported by the ML analysis (76% bootstrap support), but is more weakly supported in the MLdist analysis. The clade uniting HSP70-B and HSP70-C receives weak-to-moderate bootstrap support (63% with ML, 56% with MLdist). In both ML and MLdist analyses the (HSP70-B, HSP70-C) clade is most closely related to other euglenozoan HSP70 sequences, although support for this association is not strong (49% with ML, 64% with MLdist).

Examination of the raw alignment shows that there are at least five well-aligned sites where all (or all, except one) HSP70-B and HSP70-C sequences share a residue that is rarely or never seen in other cytosolic HSP70 sequences (four such positions are shown in Figure 2). This pattern is consistent with a close relationship between HSP70-B and HSP70-C.

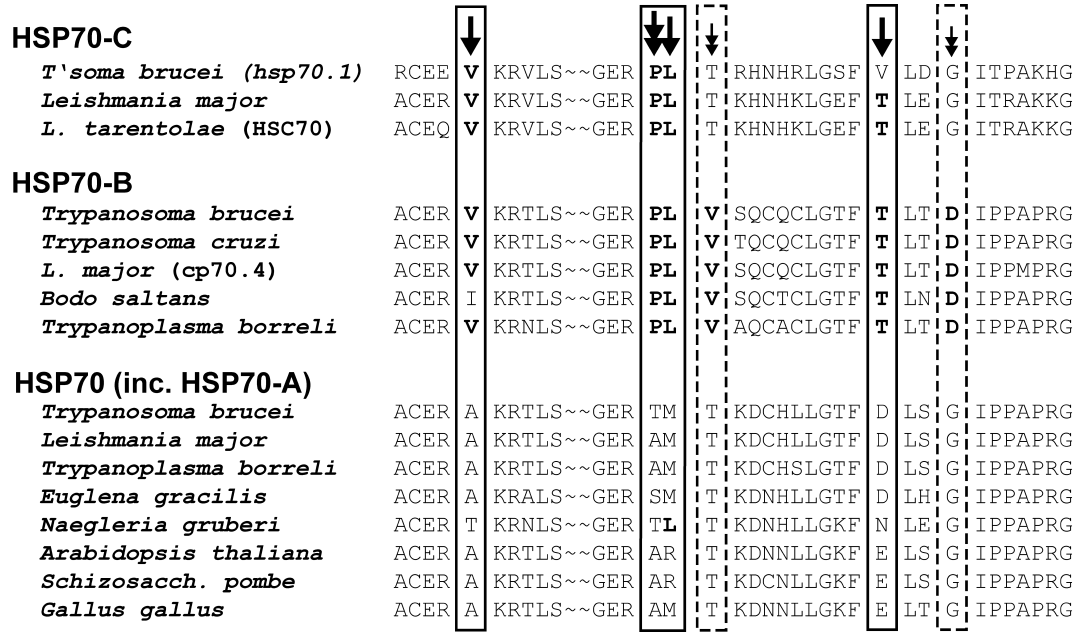

Figure 2. Portion of alignment of cytosolic HSP70 amino acid sequences, including HSP70-B and HSP70-C. Large arrows and solid boxes show sites where HSP70-B and HSP70-C share a highly conserved amino acid, to the exclusion of almost all other cytosolic HSP70 sequences, suggesting a sister group relationship. The large double arrow indicates a site where a proline residue is absolutely and uniquely conserved in HSP70-B and HSP70-C. Small double arrows and dashed boxes indicate conserved positions where HSP70-B sequences share one residue, while other cytosolic HSP70 sequences including HSP70-C share a different residue.

Furthermore, there are at least three unambiguously aligned positions where HSP70-B sequences all share one amino acid residue, but all other cytosolic HSP70 sequences, including HSP70-C, have a different residue (two such positions are shown in Fig. 2). These sites are consistent with HSP70-B forming a clade to the exclusion of HSP70-C.

In order to obtain a better estimate of the position of the putative (HSP70-B, HSP70-C) clade, we analysed data set 70AB, which does not include the highly divergent HSP70-C sequences. Both our ML and MLdist analyses place the clade, represented only by HSP70-B, as most closely related to other HSP70s from Euglenozoa (Fig. 1, right). In contrast to the 70ABC analysis, bootstrap support for this association is relatively strong (83% with ML, 75% with MLdist). Our distance analysis strongly supports the emergence of HSP70-B before the basal divergence within Euglenozoa (94% bootstrap support). However the ML tree shows HSP70-B within Euglenozoa, originating after the divergence of euglenids from diplonemids and kinetoplastids (Fig. 1, right). Bootstrap support for this position is very weak (38%), with the majority of ML bootstrap replicates (61%) supporting the sisterhood of HSP70-B and other HSP70 sequences from Euglenozoa.

Relationships amongst Kinetoplastids

In order to best estimate the organismal phylogeny of kinetoplastids we performed several protein-level analyses of HSP90 data sets (Fig. 3), HSP70 data sets with the HSP70-B and HSP70-C sequences excluded (Fig. 4) and combined HSP90/HSP70 data sets (Fig. 5). Data sets either included a diversity of other eukaryotes as outgroups (90A, 70A and CombA data sets), or included other Euglenozoa – euglenids and diplonemids – as the only outgroups (90Z, 70Z and CombZ data sets).

The basic topologies of the trees estimated from these data sets are consistent with other recent phylogenetic analysis of HSP90 and/or HSP70 proteins

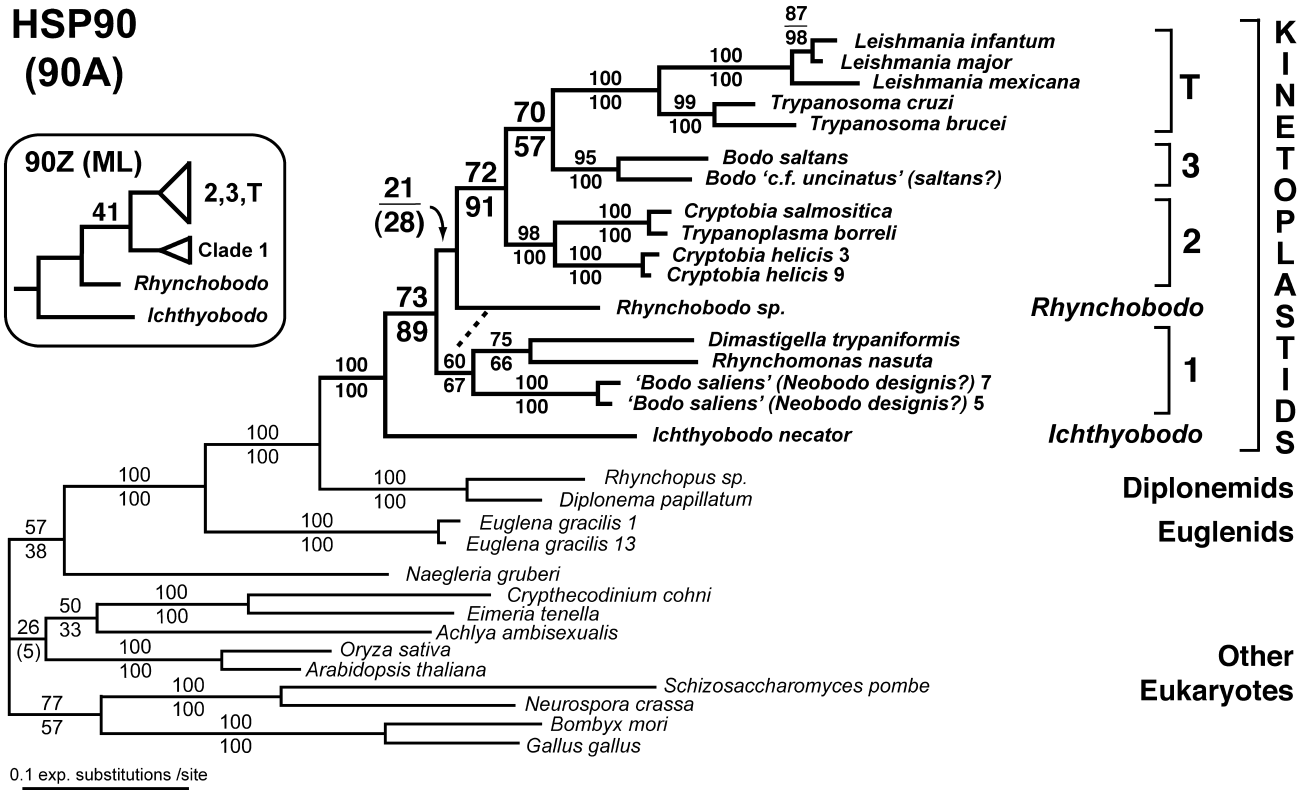

Figure 3. Maximum likelihood tree for HSP90 proteins (90A analysis). Major clades within kinetoplastids are labelled (T = trypanosomatids). Bootstrap support values for internal branches are shown (ML above, MLdist below). Support values for branches not present in the best MLdist tree are shown in parentheses. The dashed line indicates the optimal topology recovered in the MLdist analysis, where *Rhynchobodo* is sister to clade 1. Inset depicts the relationships around the base of kinetoplastids in the 90Z ML analyses, where *Rhynchobodo* is basal relative to clade 1, with the bootstrap support for this position labeled.

focused on Euglenozoa. Within a monophyletic Euglenozoa, kinetoplastids form a clade in all analyses, and branch as the sister group to diplonemids. The kinetoplastid clade generally receives strong bootstrap support (68% with 70Z, otherwise 94%+).

***Ichthyobodo*:** The new *Ichthyobodo* HSP90 sequence forms the most basal branch within kinetoplastids in all analyses (Fig. 3). Bootstrap support is fairly strong (73% with ML; 89% with MLdist) when a taxon rich outgroup is used (90A analysis). However, bootstrap support is much weaker (~45% – see Table 1) when other Euglenozoa are used as the only outgroups (90Z analysis), and trees in which *Ichthyobodo* fall elsewhere within kinetoplastids are not significantly rejected by 'Approximately Unbiased' (AU) tests of either data set (Table 2).

***Rhynchobodo* and clade 1:** In all analyses, the most basal groups within kinetoplastids other than *Ichthyobodo* are the confirmed members of clade 1 (*Rhynchomonas*, *Dimastigella* and '*Bodo saliens*' – see Simpson et al. 2002) and the new sequences from *Rhynchobodo* (Figs 3–5). In HSP90 phylogenies the confirmed members of clade 1 still form a monophyletic group, with weak-to-moderate bootstrap support (57–67%). The exact placement of *Rhynchobodo* varies between analyses. *Rhynchobodo* is sometimes placed as a separate lineage diverging immediately 'after' clade 1 (90A ML; 90Z MLdist; Fig. 3), or as the specific sister to clade 1 (90A MLdist). However, all analyses of HSP70 and combined data sets as well as the ML analysis of the 90Z data set place *Rhynchobodo* as a separate branch diverging immediately before clade 1 (Fig. 3-insert; Figs 4, 5). All analyses of HSP90 data provide very weak bootstrap support for any of the possible positions of *Rhynchobodo* (Table 1). A basal placement of *Rhynchobodo* receives moderate-to-strong bootstrap support in all analyses of the 70Z, 70A and CombZ data sets (64–82%). Bootstrap support is particularly strong in the ML analysis of CombA (83%) but is weak in the MLdist analysis (50%).

With one marginal exception, the various alternative placements for *Rhynchobodo* relative to clade 1 are not rejected by AU tests of any data set (Table 2, columns 2–4). In the cases of the HSP70 and combined data sets uncertainty in the position of the

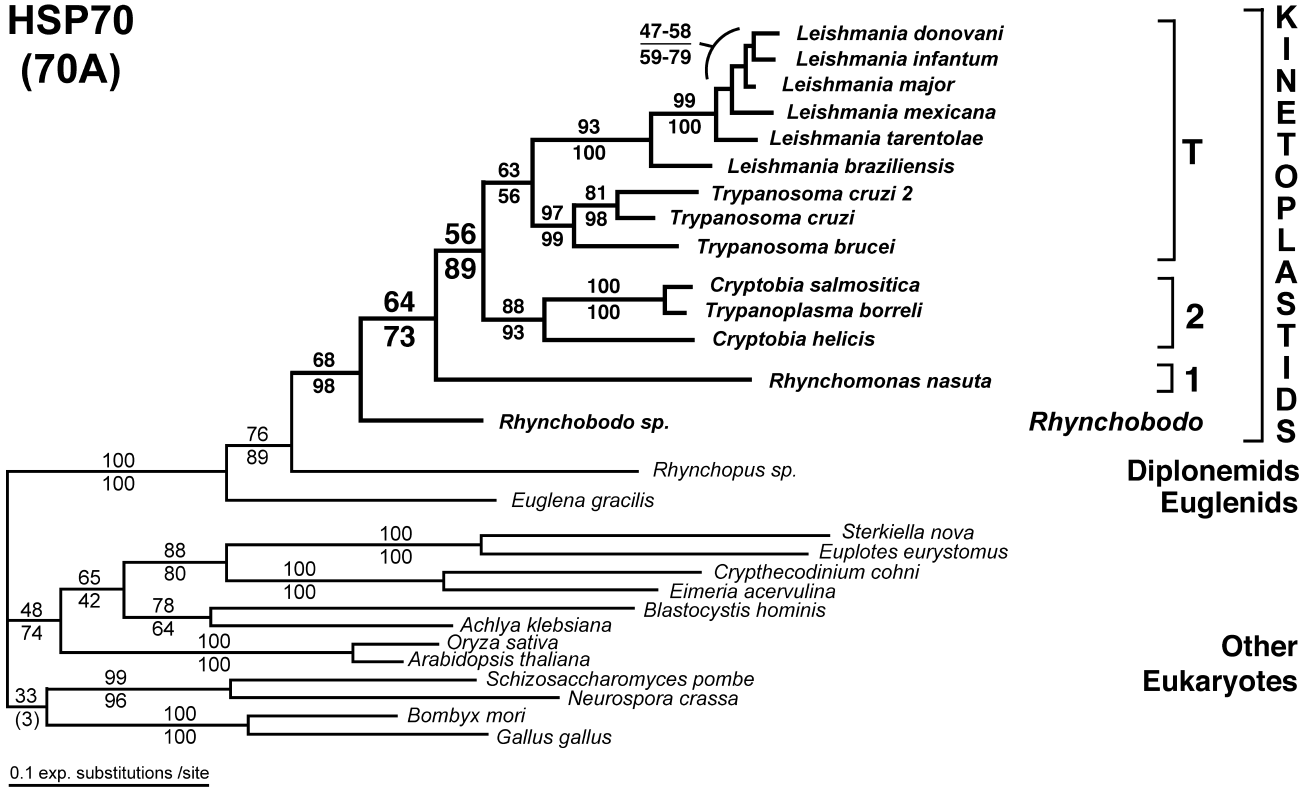

Figure 4. Maximum likelihood tree for HSP70 proteins (70A analysis). Major clades within kinetoplastids are labelled (T = trypanosomatids). Bootstrap support values for internal branches are shown (ML above, MLdist below). Some support values within *Leishmania* summarised as a single range.

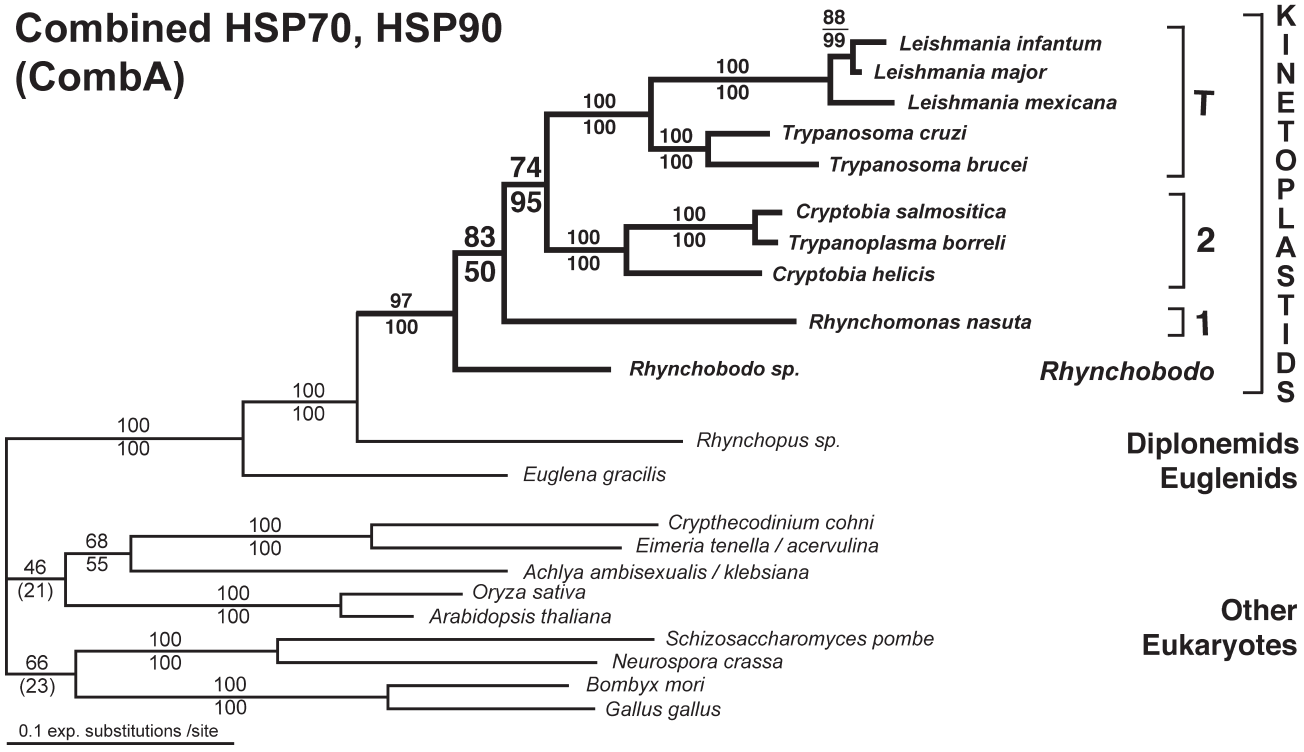

Figure 5. Maximum likelihood tree for combined HSP90 and HSP70 proteins (CombA analysis). Major clades within kinetoplastids are labelled (T = trypanosomatids). Bootstrap support values for internal branches are shown (ML above, MLdist below).

clade 1 representative *Rhynchomonas* is a complicating factor. Topologies based on the ML tree, but with *Rhynchomonas* moved all around the kinetoplastid subtree (including positions remote from the root) are the most likely alternatives to the ML tree in the set of tested trees, and confer only marginally lower likelihood on the data (ΔlnL <8 for 70A and 70Z, ΔlnL <14 for CombA and CombZ). We note that *Rhynchomonas* is a relatively long branch among studied kinetoplastids with these data sets (it is ~1.5 times the length of the *Rhynchobodo* branch in HSP70 and combined protein trees, under the evolutionary models examined here, when outgroups to kinetoplastids are excluded – data not shown)

Trypanosomatids, clade 2 and clade 3: In all analyses, the most 'apical' portion of the kinetoplastid subtree is comprised of trypanosomatids and clade 2, represented by *Cryptobia helicis*, *Trypanoplasma borreli* and *Cryptobia salmositica*. In HSP90 trees this grouping also includes bodonid clade 3 (e.g. *Bodo saltans*), for which there are no HSP70-A sequences available. The grouping of trypanosomatids and clade 2 (and clade 3) receives relatively strong bootstrap support (70%+) in all cases, except the ML analyses of the HSP70 data sets, where support is weak (Table 1). In the HSP90 analyses, clade 3 forms a specific sister group relationship with trypanosomatids, with bootstrap support between 57% and 82% (Table 1).

With all data sets, the more likely trees lacking these groupings are not rejected by AU tests (Table 2, columns 5–6), but the reason for this result differs amongst data sets. In HSP90 analyses, AU tests do not reject trees where trypanosomatids are the most basal kinetoplastids other than *Ichthyobodo*, despite these trees conferring a substantially lower likelihood on the data (>15 lnL units; Table 2, column 7). In HSP70 analyses, this basal placement of trypanosomatids is actually rejected by AU tests (Table 2, column 7), but trees in which the clade 1 representative *Rhynchomonas* is moved into the trypanosomatid-clade 2 grouping are not rejected (see above).

Clade 2 and clade 3 each receive strong bootstrap support in all analyses (83%+). The trypanosomatids always form a clade, receiving very strong bootstrap support in HSP90 and combined analyses, but only weak-to-moderate support with HSP70 (Table 1). The monophyly of trypanosomatids is not examined directly by our AU tests,

Table 1. Bootstrap support percentages for important clades. NT signifies topologies not tested by a particular analysis. Bold indicates support > 70%. Parentheses indicate topologies not present in the best tree for a particular analysis. *Ichb* = *Ichthyobodo*; *Rb* = *Rhynchobodo*; 1 = clade 1; 2 = clade 2; 3 = clade 3; T = trypanosomatids; *T'soma* = *Trypanosoma*.

	Ichb. *basal*	(*Rb*(1..T))	(1(*Rb*..T))	(*Rb*,1)	(2,3,T)	(3,T)	T	*T'soma*
90A – ML	**73**	(30)	21	(1)	**72**	**70**	**100**	**99**
90A – MLdist	**89**	(20)	(28)	29	**91**	57	**100**	**100**
90Z – ML	43	41	(10)	(13)	**81**	**82**	**100**	**100**
90Z – MLdist	46	(14)	30	(17)	**95**	69	**100**	**100**
70A – ML	N.T.	64	(21)	(4)	56	N.T.	63	**97**
70A – MLdist	N.T.	**73**	(26)	(<1)	**89**	N.T.	56	**99**
70Z – ML	N.T.	**75**	(18)	(1)	48	N.T.	58	**93**
70Z – MLdist	N.T.	**82**	(16)	(1)	**76**	N.T.	**72**	**94**
CombA – ML	N.T.	**83**	(11)	(3)	**74**	N.T.	**100**	**100**
CombA – MLdist	N.T.	50	(32)	(17)	**95**	N.T.	**100**	**100**
CombZ – ML	N.T.	**81**	(15)	(3)	**77**	N.T.	**100**	**100**
CombZ – MLdist	N.T.	68	(20)	(12)	**97**	N.T.	**100**	**100**

Table 2. Log likelihood differences ((lnL) between ML trees and specified alternative tree topologies. NT signifies topologies not tested by a particular analysis. Bold indicates topologies rejected by AU tests (* = $p < 0.05$; ** = $p < 0.01$). *Ichb* = *Ichthyobodo*; *Rb* = *Rhynchobodo*; 1 = clade 1; 2 = clade 2; 3 = clade 3; T = trypanosomatids; *T'soma* = *Trypanosoma*; para = paraphyletic.

	Ichb. not basal	*(Rb(1..T))*	*(1(Rb..T))*	(*Rb,*1)	non-(2,3,T)	non-(3,T)	T 'basal'	T grade	*T'soma* para.
90A	10.0	<0.1	best	4.6	15.8	14.2	15.8	**66.7****	**17.3***
90Z	2.9	best	2.8	5.4	17.6	15.0	17.6	**69.3****	**19.2***
70A	N.T.	best	3.8	7.2	3.1	N.T.	**19.2***	**41.9****	16.5
70Z	N.T.	best	3.0	**7.7***	2.9	N.T.	**19.3****	**36.5****	11.4
CombA	N.T.	best	9.7	13.6	9.1	N.T.	27.0	**94.9****	**31.9****
CombZ	N.T.	best	8.5	12.8	7.7	N.T.	25.8	**90.6****	**26.2****

however, all considered trees in which trypanosomatids are the paraphyletic stem group for core kinetoplastids confer markedly less likelihood on the data and are strongly rejected (Table 2, column 8). Within trypanosomatids, the taxa *Leishmania* and *Trypanosoma* always fall as two distinct clades, each with very strong bootstrap support. Paraphyly of *Trypanosoma* is not rejected by AU tests based on the HSP70 data alone, but is marginally rejected in HSP90 analyses and strongly rejected by tests of the combined data sets (Table 2, column 9).

Discussion

HSP70 Paralogs in Kinetoplastids

We can confirm the existence of at least three distinct paralogs of cytosolic HSP70 in kinetoplastids. The two 'new' divergent paralogs, which we refer to as HSP70-B and HSP70-C, each form a monophyletic group in phylogenetic analyses but appear to be most closely related to each other, rather than to other cytosolic HSP70s from Euglenozoa. While it might be thought that their association in phylogenetic trees could be a long branch attraction artefact, we suspect that this is not the case, for two reasons. Firstly, while the HSP70-C sequences are clearly the most divergent in the analysis, HSP70-B sequences are comparable to some other included HSP70s, especially the sequence from *Naegleria*. Secondly, examination of our alignments reveals several amino acid sequence signatures that HSP70-B and HSP70-C share to the exclusion of all (or almost all) other cytosolic HSP70s, most consistent with a specific relationship between HSP70-B and HSP70-C.

Several single-residue signatures unite all HSP70-B sequences to the exclusion of all other HSP70 sequences, including HSP70-C. Both this

data, and our phylogenetic trees suggest strongly that HSP70-B and HSP70-C are evolutionarily distinct clades (i.e. HSP70-C is not just the highly divergent result of a gene duplication in the recent history of HSP70-B). Given this, we can infer that their mutual divergence occurred some time ago. Although HSP70-C is known so far only from trypanosomatids, HSP70-B occurs in trypanosomatids, clade 3 and clade 2 at least. The split between HSP70-B and HSP70-C therefore dates at least to the divergence of trypanosomatids from clade 2. Obviously the divergence of the whole (HSP70-B, HSP70-C) clade from other cytosolic HSP70s must also date back this far at least. Our phylogenetic analyses consistently place these paralogs specifically with other sequences from Euglenozoa, suggesting that HSP70-B and HSP70-C are likely to be restricted to Euglenozoa. Our best estimate from maximum likelihood trees of the 70AB data set is that this divergence occurred within Euglenozoa, after the divergence of euglenids, but before the separation between studied kinetoplastids and diplonemids. However a divergence before the radiation of extant Euglenozoa receives the bulk of support from our analyses, and is therefore difficult to rule out. We predict that kinetoplastids in general will be found to have up to three distinct cytosolic HSP70s and that these may also be present in diplonemids, and perhaps euglenids.

The functions of HSP70-B and HSP70-C are poorly known. The proteins are only characterised in a couple of trypanosomatids, where they have a relatively low level of expression that is not altered significantly by heat shock (Lee et al. 1990; Searle and Smith 1993; Searle et al. 1989). HSP70-B (cp70.4) is localised to the cytoplasm in *Leishmania major* (Searle and Smith 1993). In *Leishmania tarentolae* HSP70-C is involved in antimony tolerance, but so is HSP70-A (Brochu et al. 2004). The recognition that there are multiple deeply distinct paralogs of HSP70 in some, perhaps all, kinetoplastids should spur interest in determining their various roles in these cells.

Position of *Ichthyobodo*

Our HSP90 trees support an early diverging position of *Ichthyobodo* within kinetoplastids, consistent with several recent analyses of SSUrRNA data (Callahan et al. 2002; Dyková et al. 2003; von der Heyden et al. 2004; López-Garcia et al. 2003; Moreira et al. 2004). *Ichthyobodo*, together with symbionts of amoebae assignable to (or at least similar to) *Perkinsiella* form a lineage apart from all other studied kinetoplastids, including trypanosomatids and the 'bodonids' assigned to clades 1, 2 and 3, as well as *Rhynchobodo* sp. (see below). Moreira et al. (Moreira et al. 2004) have recently assigned this 'core kinetoplastid clade' a formal name: Metakinetoplastina. Our present study provides further evidence that this taxon is a natural one, and confirms that *Ichthyobodo* holds an important position for understanding the early evolution of kinetoplastids.

Rhynchobodo and Clade 1

Despite inclusion in both HSP90 and HSP70 datasets, the exact evolutionary position of *Rhynchobodo* remains an intriguing problem. In SSUrRNA analyses, *Rhynchobodo* is usually closely associated with other members of clade 1 within the core kinetoplastid subtree, either attaching to the base of clade 1, or falling 'within' clade 1 as a specific relative of *Dimastigella* and *Rhynchomonas* (von der Heyden et al. 2004; Moreira et al. 2004; Nikolaev et al. 2003; Simpson et al. 2002). Either placement is consistent with an assignation of *Rhynchobodo* to clade 1. However the majority of our protein analyses place *Rhynchobodo* as a separate lineage from clade 1, placing the root of the core kinetoplastid subtree between these two groups, usually with *Rhynchobodo* being the more basal of the two.

We are reluctant to take our majority result at face value. We note that the statistical support for the basal position of *Rhynchobodo* within core kinetoplastids is often weak, and tends to be strongest with the data sets with the weakest ingroup taxon sampling – those based on the HSP70 and combined protein alignments. *Rhynchomonas*, the sole confirmed clade 1 representative, is difficult to place in analyses of these data sets. It is a relatively long branch that can be moved to widely differing positions within the kinetoplastid subtree with comparatively little impact on the likelihood of the data. The uncertainty in the placement of *Rhynchobodo* might be better viewed as uncertainty in the placement of *Rhynchomonas*, at least with these data sets. It is curious, however, that this problematic *Rhynchomonas* sequence should have weaker affinity for the base of the kinetoplastid subtree than the less divergent *Rhynchobodo* sequence – this being the opposite pattern to that expected from classical subtree mis-rooting by long branch attraction to outgroups (e.g. Philippe and Adoutte 1998). The addition of more (and hopefully less divergent) clade 1 representatives to the HSP70 database could be key to a better resolution of the phylogenetic position of *Rhynchobodo*.

In summary, the phylogenetic position of *Rhynchobodo* remains unresolved, with a placement of

this taxon separate from clade 1 being more plausible than before. The recent creation of a formal taxon Neobodonidae including both confirmed 'clade 1' organisms and *Rhynchobodo* (Moreira et al. 2004) would seem to be premature.

The Monophyly of *Trypanosoma*

Early analysis of ribosomal RNA sequences called into question the monophyly of the *Trypanosoma,* placing members of this taxon as a basal grade for all trypanosomatids (Gomez et al. 1991; Landweber and Gilbert 1994; Maslov et al. 1994). Later phylogenetic analyses of ribosomal RNA sequences with improved taxon sampling and methods suggested instead that *Trypanosoma* was monophyletic (Lukeš et al. 1997; Stevens et al. 1999; Wright et al. 1999). The phylogenetic trees of the great majority of available protein data sets also support the monophyly of *Trypanosoma* (Alvarez et al. 1996; Gaziová and Lukeš 2003; Hannaert et al. 1998; Hughes and Piontkivska 2003b; Qian and Keeling 2001; Wiemer et al. 1995). However Hughes and Piontkivska (Hughes and Piontkivska 2003a, b) recently revived the idea of *Trypanosoma* paraphyly (or polyphyly), based on their analyses of a trypanosomatid-rich SSUrRNA data set. While noting that the bulk of protein trees support the monophyly of *Trypanosoma*, they consider the question open, citing the poor taxon sampling of all protein data sets as a reason for skepticism as to their phylogenetic accuracy (Hughes and Piontkivska 2003b).

Hughes and Piontkivska (Hughes and Piontkivska 2003a, b) clearly consider the critical issue to be taxon sampling within trypanosomatids. However, as distinguishing between monophyly and paraphyly of *Trypanosoma* can be viewed as an issue of where to root the trypanosomatid subtree, taxon sampling among possible/probable close outgroups to trypanosomatids could be at least as important a concern. While remaining inferior to some proteins with respect to within-trypanosomatid taxon sampling, our heat shock protein data sets have the best sampling available for other Euglenozoa. Our trees all recover *Trypanosoma* as a clade, with very strong bootstrap support in the case of HSP90 and combined protein trees, while trypanosomatid paraphyly is strongly rejected in AU tests of our combined protein data set. Improving the taxon sampling of these protein data sets, has, if anything, further strengthened the support for monophyly of *Trypanosoma*, not weakened it. The evidence for *Trypanosoma* monophyly remains stronger by far than the case for non-monophyly.

'Trypanosomatids late'

Determining the phylogenetic position of trypanosomatids within kinetoplastids is central to understanding the evolution of their biological attributes, including parasitism, patterns of kinetoplast DNA organization and RNA editing. For at least the last two decades there has been a widespread belief that trypanosomatids are descended in some way from typical 'bodonid' ancestors (Kivic and Walne 1984; Vickerman 1990). This general proposal was based originally on morphological considerations, but received further support in the late nineteen nineties from the first bodonid-rich molecular phylogenetic analyses (Atkins et al. 2000; Blom et al. 1998; Doležel et al. 2000; Lukeš et al. 1997; Wright et al. 1999). However, recent re-analyses of SSUrRNA data are leading to the consideration of 'trypanosomatids early' proposals, in which trypanosomatids would be the sister to (or possibly ancestral group for) almost all bodonid kinetoplastids (von der Heyden et al. 2004; Hughes and Piontkivska 2003a; Moreira et al. 2004).

Previous heat shock protein phylogenies generally support the 'late' emergence of trypanosomatids within (core) kinetoplastids. Analyses of both HSP70 and HSP90 proteins that focus on Euglenozoa have supported a clade consisting of trypanosomatids, and clade 2, and where taxon sampling is available, clade 3. However, statistical support for this clade varies with different combinations of sequence data, taxon sampling and analysis methods (Simpson and Roger 2004; Simpson et al. 2002), and has tended to be weak in the analyses with the strongest methodological basis and taxon sampling. In particular, previous ML analyses of HSP90 that focus on Euglenozoa provide $<55\%$ bootstrap support for both the (trypanosomatid, clade 3, clade 2) grouping and the (trypanosomatid, clade 3) cluster when more than a single outgroup is employed (Simpson and Roger 2004; Simpson et al. 2002). Furthermore, a recent study of broader eukaryote phylogeny using HSP90 protein sequences, weakly but consistently places trypanosomatids as the earliest diverging clade within kinetoplastids (Stechmann and Cavalier-Smith 2003). This latter study examines less of the length of the HSP90 coding sequence than did Simpson et coll., but includes more and different outgroups to kinetoplastids.

Our present study of heat shock protein markers employs an improved taxon sampling within kinetoplastids. All of our analyses group trypanosomatids with clade 2 (and clade 3) bodonids to the exclusion of both clade 1 and *Rhynchobodo*. With most combinations of data sets and methods this grouping re-

ceives relatively strong support. In particular, bootstrap support for this clade is 70%+ in ML analyses of the HSP90 data, irrespective of which outgroup is employed. This level of support in a methodologically strong analysis with good taxon sampling in both ingroup and outgroup constitutes an important novel result. Our analysis thus provides substantially improved positive support for the 'late' emergence of trypanosomatids within kinetoplastids, and is the strongest and most reliable phylogenetic evidence available in favour of this proposal. Our ML analyses of HSP90 data also provide improved support for a placement of trypanosomatids as the specific sister group to bodonid clade 3, strengthening the hypothesis that *Bodo saltans* is a particularly close relative of trypanosomatids (Blom et al. 1998; Simpson et al. 2002).

By contrast, several recent analyses of SSUrRNA data find some support for an 'early' divergence of trypanosomatids. Hughes and Piontkivska (Hughes and Piontkivska 2003a, b) examine a data set with a very strong taxon sampling of trypanosomatids, but with two euglenids as the only outgroups. They recover trypanosomatids as a paraphyletic stem group for (core) kinetoplastids. Moreira et al. (Moreira et al. 2004), and von der Heyden et al. (von der Heyden et al. 2004) consider a more even sampling across core kinetoplastids and, critically, include several sequences that break the basal branch leading to core kinetoplastids. Many of their analyses suggest that trypanosomatids as a clade could form the earliest diverging branch within core kinetoplastids.

We have an alternative interpretation of the results seen by Hughes and Piontkivska (Hughes and Piontkivska 2003a, b). As they employ a taxonomically restricted and evolutionarily divergent outgroup (two members of Euglenales), the basal branch leading to kinetoplastids is surely extremely long in their analyses. Further, the number of included sites is relatively high (1431 – Hughes and Piontkivska 2003b), implying that many rapidly evolving sites are included that would tend to be excluded from most similar analyses (Doležel et al. 2000; Moreira et al. 2004; Simpson and Roger 2004; Simpson et al. 2002). These features suggest that substitutional saturation of comparisons between kinetoplastids and outgroups would be even more severe than in previous analyses (Busse and Preisfeld 2002; Moreira et al. 2004), and the placement of the root perhaps even less trustworthy. The placement of the root within trypanosomatids in these analyses (in several different places, depending on the methods employed, and usually with weak support) could easily be a simple consequence of including many more trypanosomatids than bodonids in the analyses, providing a better chance for a root placed 'at random' to attach somewhere within the former group. It is notable that a rooting of core kinetoplastids within trypanosomatids is strongly rejected by AU tests of all of our heat shock protein data sets (Table 2). Also, there are now several protein data sets where sequences of a reasonable length are available for a single bodonid, in addition to several trypanosomatids. These include cytochrome *c* oxidase subunits I and II, apocytochrome b, Glyceraldehyde-3-Phosphate Dehydrogenase (GAPDH), and topoisomerase II. In each case available phylogenies consistently place trypanosomatids as a clade to the exclusion of the bodonid, usually with strong statistical support (Gaziová and Lukeš 2003; Hannaert et al. 1998; Lukeš et al. 1994; Maslov et al. 1999; Qian and Keeling 2001; Wiemer et al. 1995).

The recent analyses by Moreira et al. (Moreira et al. 2004) and von der Heyden et al. (von der Heyden et al. 2004) explicitly aim to minimise the evolutionary distance between core kinetoplastids and other taxa, by including new closer outgroups, especially *Ichthyobodo* and related sequences (see also López-Garcia et al. 2003). The basal position of trypanosomatids is recovered with many, but not all, combinations of taxon inclusion and methods (generally speaking, the most sophisticated). However, the topology always garners weak statistical support, even in analyses employing relatively sophisticated evolutionary models, for example, 53% ML bootstrap support, 63% or <50% ML distance bootstrap support (across several analyses) and 0.74 bayesean posterior probability (von der Heyden et al. 2004; López-Garcia et al. 2003; Moreira et al. 2004). We note that some other analyses employing some of these new outgroups also do not recover a basal placement of trypanosomatids (Callahan et al. 2002; Dyková et al. 2003). Interestingly, however, only in analyses of HSP70 data sets do our AU tests reject all 'trypanosomatids early' topologies in which trypanosomatids are a clade.

In summary, the evidence for a late origin of trypanosomatids from heat shock protein phylogenies ('almost always recovered in analyses of two different molecular markers, with statistical support ranging from very strong to weak') is currently much more compelling than the evidence for an early origin of trypanosomatids from well-sampled SSUrRNA trees ('often recovered in analyses of a single molecular marker, with statistical support ranging from moderate to very weak'). Efforts to further improve the taxon sampling of protein data sets would be highly desirable. Nonetheless, we suspect that

nuclear-encoded protein phylogenies are already providing a more consistent and accurate estimate of the deep topology of the core kinetoplastid subtree than are the SSUrRNA data, and may represent the more efficient avenue to further insight into this important portion of the tree of life.

Methods

Identity of *Rhynchobodo* sp.: *Rhynchobodo* sp. ATCC 50359 was grown in 802 media enriched with *Klebsiella*. Live cells were examined by differential interference contrast light microscopy, using a Zeiss Axioplan II microscope with an Axiocam HR digital camera.

Cells are around 8–12 μm long with a substantial apical rostrum comprising more than one third of the length of the cell (Figs 6a, b). Two flagella insert into a distinct forward-directed flagellar pocket (Fig. 6b). Both flagella are thick; the anterior flagellum is slightly longer than the cell, and the posterior flagellum is slightly more than twice the length of the cell (Fig 6a–c). Cells swim rapidly rather than gliding, and ingest the co-cultured bacteria. Cells are normally rigid, but can undergo active squirming under stress. Swimming cells may give the impression of possessing a subtle spiral groove. This isolate appears very similar to other accounts of small *Rhynchobodo* spp. (see Brugerolle 1985; Patterson and Simpson 1996), supporting the assignment of ATCC 50359 to *Rhynchobodo* (see Moreira et al. 2004).

Gene amplification and sequencing: Genomic DNA from *Rhynchobodo* sp. ATCC 50359 was extracted using a CTAB protocol (Clark and Diamond 1991).

Genomic DNA from *Ichthyobodo necator* (Henneguy 1883) Pinto 1928 (striped bass isolate), *Cryptobia helicis* Leidy 1846, *Cryptobia salmositica* Katz 1951, *Trypanoplasma borreli* Laveran and Mesnil 1901 (strain Tt-JH), and *Bodo saltans* Ehrenberg 1830 (Konstanz isolate) was extracted as described previously (Callahan et al. 2002; Doležel et al. 2000).

Near-complete coding sequences for cytosolic *hsp90* from *Ichthyobodo necator* and *Rhynchobodo* sp. were amplified by PCR using primers 100X and 970X (Simpson et al. 2002). Near-complete coding sequences for cytosolic *hsp70* (the gene for HSP70-A) were isolated from *Rhynchobodo* sp., *Cryptobia helicis* and *Cryptobia salmositica* using primers 65CE and GGMP2 (Simpson and Roger 2004). A near-complete coding sequence for the gene for HSP70-B was isolated from *Bodo saltans* using 65CE and GGMP2, with the latter oligonucleotide mispriming almost immediately downstream of the terminal stop codon. A partial (~70% complete) coding sequence for this gene was subsequently amplified from *Trypanoplasma borreli* using 65CE

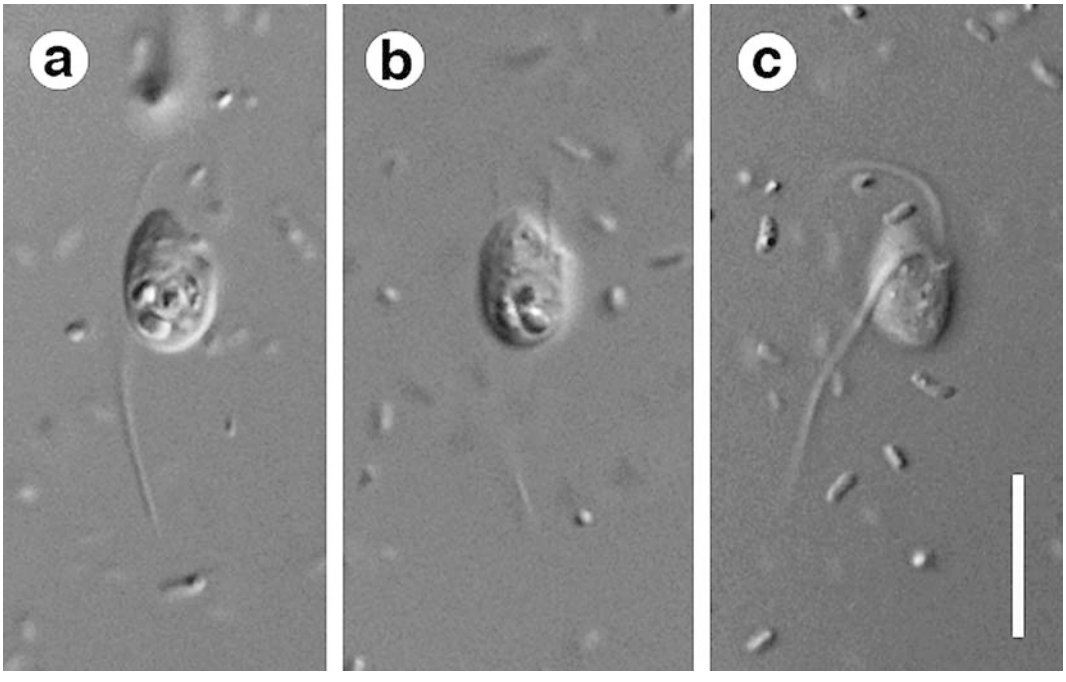

Figure 6. Light micrographs of *Rhynchobodo* sp. ATCC 50239 (Cells slightly compressed for observation) **a.** Cell profile. **b.** View showing insertion of flagella into pocket. **c.** Detail of emergent flagella. Scale bar represents 10 μm.

and an HSP70-B-specific reverse primer, 602P (designed against the amino acid sequence NTDGILIV with the 5′-3′ DNA sequence ACGATGAGGATNCC RTCNGTRTT). Annealing temperatures were 50–55 °C.

All amplicons were gel-purified, then cloned into a TA vector (TOPO 2.1. Invitrogen). One to four clones were partially sequenced, with one clone selected for complete bi-directional sequencing. Limited, mostly silent (synonymous) heterogeneity was detected among clones. No introns were detected in any of the new sequences. The new sequences have the Genbank accession numbers AY651251–AY651257.

Data set construction: Conceptual amino acid translations were aligned by eye with homologous sequences from other eukaryotes. Some relevant sequences from ongoing genome projects on trypanosomatids were identified through BLAST searches and were also added to the alignment as amino acids. Cytosolic *hsp90*, and the genes for HSP70-A and HSP70-C were identified in genomic data from *Leishmania major*. Genes for HSP70-B were identified in genomic data from *Trypanosoma brucei* and *Trypanosoma cruzi*. As the original Genbank entry for *Leishmania major hsp70.4* (i.e. the gene for HSP70-B) contained multiple frame-shift sequencing errors, the protein sequence was inferred from genomic data. For *Trypanosoma cruzi* and *Trypanosoma brucei* (chromosome 7), preliminary sequence data was obtained from The Institute for Genomic Research website at http://www.tigr.org. Sequencing of *Trypanosoma brucei* chromosome 7 is supported by an award from the National Institute of Allergy and Infectious Diseases, National Institutes of Health. Sequence data for *Leishmania major* chromosomes 26, 28 and 32 was obtained from The Sanger Institute website at http://www.sanger.ac.uk/Projects/L_major/. Sequencing of *Leishmania major* chromosomes 26, 28 and 32 was accomplished as part of the *Leishmania* Genome Network with support by The Wellcome Trust.

For HSP70 analyses, sequences from Euglenozoa were considered together with 12 relatively short-branching full-length sequences representing other eukaryotic groups, and the more divergent sequence available for the heteroloboseid *Naegleria gruberi* (Simpson and Roger 2004). Several data sets were constructed that differed both in site inclusion and sequence representation. In all cases, sites were retained from downstream of the 65CE site only. Three considered sequences included missing data: The partial sequences for *Euglena gracilis* HSP70 and *Trypanoplasma borreli* HSP70-B were missing some sites from the C-terminal end of the alignment (55–131 analysed sites, depending on the data set). The gene for HSP70-C from *Trypanosoma brucei* (*hsp70.1*) has been involved in gene conversion with a nearby cluster of five genes for HSP70-A (*hsp70.2–6*), such that at least three tracts of *hsp70.1* and *hsp70.2–6* are identical at the nucleotide level (Lee et al. 1990). Our full protein alignments indicate that *hsp70.1* was converted to conform to the *hsp70.2–6* sequence, rather than vice versa. The regions of *Trypanosoma brucei* HSP70-C corresponding to the three converted tracts were therefore re-coded as missing data, affecting 103 analysed sites.

The broadest HSP70 data set, ‘70ABC’, contains diverse cytosolic HSP70s including both HSP70-B and HSP70-C, and 482 sites (to reduce computational effort, three *Leishmania* HSP70-A sequences were omitted). Data set ‘70AB’ excludes the more divergent HSP70-C and retains 573 sites. The remaining data sets exclude all HSP70-B and HSP70-C sequences, and HSP70 from *Naegleria*. Data set ‘70A’ includes all HSP70-A sequences from Euglenozoa, plus all outgroups, retaining 577 sites. Data set ‘70Z’ includes only Euglenozoa and 592 sites.

For phylogenetic analyses of HSP90, sequences from all groups of Euglenozoa were considered, plus 10 full-length and relatively short-branching sequences representing other eukaryotic groups. The database of kinetoplastid HSP90s contains sequences from isolates originally labelled as *Bodo saliens* and *Bodo uncinatus*. SSUrRNA trees show that these isolates are nested within clades predominantly comprising isolates identified as the morphospecies *Neobodo designis* Vickerman 2004 (bas. *Bodo designis* Skuja 1948) and *Bodo saltans* Ehrenberg 1830 respectively (von der Heyden et al. 2004), suggesting that *Neobodo designis* and *Bodo saltans* are the correct morphospecies identifications for these isolates (see Fig. 3). A total of 605 sites between the 100X and 970X priming sites were retained for analysis, as per Simpson et al. (Simpson et al. 2002). Two data sets were constructed that differed only in taxon representation – 90A including all considered taxa, while 90Z including only Euglenozoa.

Combined HSP70/HSP90 data sets were constructed by concatenating the HSP90 and HSP70 alignments (excluding both HSP70-B and HSP70-C). Named species were used as OTUs, with two exceptions: *Eimeria* (*Eimeria tenella* HSP90 combined with *Eimeria acervulina* HSP70) and *Achlya* (*Achlya ambisexualis* HSP90 and *Achlya klebsiana* HSP70). Other species-taxa included in only one of

the two original alignments were excluded, as was *Naegleria gruberi*. Where multiple sequences were available for one nominal species, only the shortest branching sequence was used. Data set 'CombA', including both Euglenozoa and outgroups, retained 1183 sites, while data set 'CombZ' includes Euglenozoa only and retained 1205 sites.

Phylogenetic analysis: All data sets were analysed by Maximum Likelihood (ML) using PMBML 3.6a2.1 (E. Tillier, pers. comm., Felsenstein 2000), employing the JTT amino acid substitution matrix, with among-site rate variation modeled as a α distribution approximated by eight equiprobable discrete rate categories. These rates were estimated from a neighbour-joining tree using TREE-PUZZLE 5.0 (Strimmer and von Haeseler 1996), and then were inputted manually into PMBML. The values of the α shape parameter were as follows: 70ABC – 0.58; 70AB – 0.41; 90A – 0.54; 90Z – 0.40; 70A – 0.37; 70Z – 0.30; CombA – 0.44; CombZ – 0.37. The ML trees were searched for using 10 random taxon addition sequences, each followed by 'global rearrangements'. Bootstrap analyses were performed using the same settings, but with one taxon addition sequence per replicate. The numbers of ML bootstrap replicates examined were as follows: 70ABC – 100; 70AB – 100; 90A – 200; 90Z – 200; 70A – 200; 70Z – 400; CombA – 200; CombZ – 400.

All data sets were also analysed using a maximum likelihood distance (MLdist) method, employing the WAG substitution matrix, with among-site rate variation modeled by an eight category discrete approximation of a Γ distribution (values of α were very similar to those estimated with the JTT matrix for the ML analysis –see above). Pairwise maximum likelihood estimates of evolutionary distances were calculated using TREE-PUZZLE, and the best least-squares tree was found using FITCH 3.6a (Felsenstein 2000), with 50 random addition sequences and global rearrangements. Bootstrap analyses with 2000 replicates were performed using PUZZLEBOOT 1.03 (http://hades.biochem.dal.ca/Rogerlab/Software/software.html) and FITCH (3–5 random taxon additions with rearrangements per replicate).

Alternative topology tests: Our ML trees were compared to sets of other plausible trees and topologies of particular interest using the 'approximately unbiased' (AU) test of Shimodaira (Shimodaira 2002). The test sets included trees that were similar to the ML topology, but showed i) alternative reconstructions of kinetoplastid relationships where *Ichthyobodo* was not basal (90A and 90Z only), ii) permutations of the relationships amongst major groups of core kinetoplastids, iii) topologies where *Trypanosoma* was not monophyletic, and iv) topologies where trypanosomatids gave rise to core kinetoplastids. The 90A and 90Z test sets included 65 trees. The test sets for 70A, 70Z, CombA and CombZ included 31 trees (test trees, and selection method are available by request). For all tested trees, site likelihoods were calculated in PAML 3.13a (Yang 1997) using the same model as the ML tree search. Approximately unbiased tests were then performed using CONSEL v. 0.1 (Shimodaira and Hasegawa 2001), with default parameters.

Acknowledgements

The authors thank J. Lukeš, S. von der Heyden and an anonymous reviewer for critical comments, K. Vickerman and A. Stechmann for discussion, and D. Moreira, S. von der Heyden and C. Brochu for sharing results prior to publication. AGBS and AJR thank the Canadian Institute for Advanced Research, Program in Evolutionary Biology for fellowship support. This research was supported in part by NSERC Grant 227085-00 awarded to AJR. For *Trypanosoma cruzi*, and *Trypanosoma brucei* (chromosome 7), preliminary sequence data was obtained from The Institute for Genomic Research website at http://www.tigr.org. Sequencing of *Trypanosoma brucei* chromosome 7 is supported by an award from the National Institute of Allergy and Infectious Diseases, National Institutes of Health. Sequencing of *Trypanosoma cruzi* was funded by the National Institute of Allergy and Infectious Disease (NIAID). Sequence data for *Leishmania major* chromosomes 26, 28 and 32 was obtained from The Sanger Institute website at http://www.sanger.ac.uk/Projects/L_major/. Sequencing of *Leishmania major* chromosomes 26, 28 and 32 was accomplished as part of the Leishmania Genome Network with support by The Wellcome Trust.

References

Alvarez F, Cortinas MN, Musto H (1996) The analysis of protein coding genes suggests monophyly of *Trypanosoma*. Mol Phylogenet Evol **5**: 333–343

Arndt H, Dietrich D, Auer B, Cleven E, Grafenham T, Weitere M, Mylnikov AP (2000) Functional Diversity of Heterotrophic Flagellates in Aquatic Ecosystems. In Leadbeater BSC, Green JC (eds) The flagellates. Taylor and Francis, London, pp 240–268

Atkins MS, Teske AP, Anderson OR (2000) A survey of flagellate diversity at four deep-sea hydrothermal vents in the eastern Pacific Ocean using structural and molecular approaches. J Eukaryot Microbiol **47**: 400–411

Blom D, de Haan A, van den Berg M, Sloof P, Jirku M, Lukeš J, Benne R (1998) RNA editing in the free-living bodonid *Bodo saltans*. Nucleic Acids Res **26**: 1205–1213

Brochu C, Haimeur A, Ouellette M (2004) The heat shock protein HSP70 and heat shock cognate protein HSC70 contribute to antimony tolerance in the protozoan parasite *Leishmania*. Cell Stress Chaperon **9**: 294–303

Brugerolle G (1985) Des trichocystes chez les bodonides, un caractère phylogénétique supplémentaire entre Kinetoplastida et Euglenida. Protistologica **21**: 339–348

Busse I, Preisfeld A (2002) Phylogenetic position of *Rhynchopus* sp. and *Diplonema ambulator* as indicated by analyses of euglenozoan small subunit ribosomal DNA. Gene **284**: 83–91

Callahan H, Litaker RW, Noga EJ (2002) Molecular taxonomy of the suborder Bodonina (Order Kinetoplastida), including the important fish parasite, *Ichthyobodo necator*. J Eukaryot Microbiol **49**: 119–128

Clark CG, Diamond LS (1991) The Laredo strain and other '*Entamoeba histolytica*-like' amoebae are *Entamoeba moshkovskii*. Mol Biochem Parasitol **46**: 11–18

Doležel D, Jirku M, Maslov DA, Lukeš J (2000) Phylogeny of the bodonid flagellates (Kinetoplastida) based on small-subunit rRNA gene sequences. Int J Syst Evol Microbiol **50**: 1943–1951

Donelson JE, Gardner MJ, El-Sayed NM (1999) More surprises from Kinetoplastida. Proc Natl Acad Sci USA **96**: 2579–2581

Dyková I, Fiala I, Lom J, Lukeš J (2003) *Perkinsiella amoebae*-like endosymbionts of *Neoparamoeba* spp., relatives of the kinetoplastid *Ichthyobodo*. Europ J Protistol **39**: 37–52

Felsenstein J (2000) PHYLIP (Phylogeny inference package), version 3.6a. Distributed by the author, University of Washington, Seattle

Gaziová I, Lukeš J (2003) Mitochondrial and nuclear localization of topoisomerase II in the flagellate *Bodo saltans* (Kinetoplastida), a species with non-catenated kinetoplast DNA. J Biol Chem **278**: 10900–10907

Gomez E, Valdés AM, Piñero D, Hernandez R (1991) What is a genus in the trypanosomatidae family? Phylogenetic analysis of two small rRNA. Mol Biol Evol **8**: 254–259

Hannaert V, Opperdoes FR, Michels PA (1998) Comparison and evolutionary analysis of the glycosomal glyceraldehyde-3-phosphate dehydrogenase from different Kinetoplastida. J Mol Evol **47**: 728–738

von der Heyden S, Chao EE, Vickerman K, Cavalier-Smith T (2004) Ribosomal RNA phylogeny of bodonid and diplonemid flagellates and the evolution of Euglenozoa. J Eukaryot Microbiol **51**: 402–416

Hollar L, Lukeš J, Maslov DA (1998) Monophyly of endosymbiont containing trypanosomatids: phylogeny versus taxonomy. J Eukaryot Microbiol **45**: 293–297

Hughes AL, Piontkivska H (2003a) Phylogeny of Trypanosomatidae and Bodonidae (Kinetoplastida) based on 18S rRNA: Evidence for paraphyly of *Trypanosoma* and six other genera. Mol Biol Evol **20**: 644–652

Hughes AL, Piontkivska H (2003b) Molecular phylogenetics of Trypanosomatidae: contrasting results from 18S rRNA and protein phylogenies. Kinetoplastid Biol Dis **2**: Article 15

Kivic PA, Walne PL (1984) An evaluation of a possible phylogenetic relationship between the Euglenophyta and Kinetoplastida. Orig Life **13**: 269–288

Landweber LF, Gilbert W (1994) Phylogenetic analysis of RNA editing: a primitive genetic phenomenon. Proc Natl Acad Sci USA **91**: 918–921

Lee MG, Polvere RI, Van der Ploeg LH (1990) Evidence for segmental gene conversion between a cognate *hsp70* gene and the temperature-sensitively transcribed *hsp70* genes of *Trypanosoma brucei*. Mol Biochem Parasitol **41**: 213–220

López-Garcia P, Philippe H, Gail F, Moreira D (2003) Autochthonous eukaryotic diversity in hydrothermal sediment and experimental microcolonizers at the Mid-Atlantic Ridge. Proc Natl Acad Sci USA **100**: 697–702

Lukeš J, Guilbride DL, Votypka J, Zikova A, Benne R, Englund PT (2002) Kinetoplast DNA network: evolution of an improbable structure. Euk Cell **1**: 495–502

Lukeš J, Jirku M, Doležel D, Kral'ová I, Hollar L, Maslov DA (1997) Analysis of ribosomal RNA genes suggests that trypanosomes are monophyletic. J Mol Evol **44**: 521–527

Lukeš J, Arts GJ, van den Berg M, de Haan A, Opperdoes FR, Sloof P, Benne R (1994) Novel pattern of editing regions in mitochondrial transcripts of the cryptobiid *Trypanoplasma borreli*. EMBO J **13**: 5086–5098

Marin B, Palm A, Klingberg M, Melkonian M (2003) Phylogeny and taxonomic revision of plastid-containing euglenophytes based on SSU rDNA sequence comparisons and synapomorphic signatures in the SSU rRNA secondary structure. Protist **154**: 99–145

Maslov DA, Yasuhira S, Simpson L (1999) Phylogenetic affinities of *Diplonema* within the Euglenozoa as inferred from the SSU rRNA gene and partial CO1 protein sequences. Protist **150**: 33–42

Maslov DA, Avila HA, Lake JA, Simpson L (1994) Evolution of RNA editing in kinetoplastid protozoa. Nature **368**: 345–348

Merzlyak E, Yurchenko V, Kolesnikov AA, Alexandrov K, Podlipev SA, Maslov DA (2001) Diversity and phylogeny of insect trypanosomatids based on small subunit rRNA genes: polyphyly of *Leptomonas* and *Blastocrithidia*. J Eukaryot Microbiol **48**: 161–169

Moreira D, Lopez-Garcia P, Vickerman K (2004) An updated view of kinetoplastid phylogeny using environmental sequences and a closer outgroup: Proposal for a new classification of the Class Kinetoplastea. Int J Syst Evol Microbiol **54**: 1861–1875

Nikolaev SI, Mylnikov AP, Berney C, Fahrni JF, Petrov NB, Pawlowski J (2003) The taxonomic position of *Klosteria bodomorphis* gen. and sp. nov. (Kinetoplastida) based on ultrastructure and SSU rRNA gene sequence analysis. Protistology **3**: 126–135

Patterson DJ, Simpson AGB (1996) Heterotrophic flagellates from coastal marine and hypersaline sediments in Western Australia. Europ J Protistol **32**: 423–448

Philippe H, Adoutte A (1998) The Molecular Phylogeny of Eukaryota: Solid Facts and Uncertainties. In Coombs GH, Vickerman K, Sleigh MA, Warren A (eds) Evolutionary Relationships among Protozoa. Chapman and Hall, London, pp 25–56

Qian Q, Keeling PJ (2001) Diplonemid Glyceraldehyde-3-Phosphate Dehydrogenase (GAPDH) and prokaryote-to-eukaryote lateral gene transfer. Protist **152**: 193–201

Searle S, Smith DF (1993) *Leishmania major*: characterisation and expression of a cytoplasmic stress-related protein. Exp Parasitol **77**: 43–52

Searle S, Campos A, Coulson R, Spithill T, Smith D (1989) A family of heat shock protein 70–related genes are expressed in the promastigotes of *Leishmania major*. Nucleic Acids Res **17**: 5081–5095

Shimodaira H (2002) An approximately unbiased test of phylogenetic tree selection. Syst Biol **51**: 492–508

Shimodaira H, Hasegawa M (2001) CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics **17**: 1246–1247

Simpson AGB, Roger AJ (2004) Protein phylogenies robustly resolve the deep-level relationships within Euglenozoa. Mol Phylogenet Evol **30**: 201–212

Simpson AGB, Lukeš J, Roger AJ (2002) The evolutionary history of kinetoplastids and their kinetoplasts. Mol Biol Evol **19**: 2071–2083

Simpson L, Thiemann OH, Savill NJ, Alfonzo JD, Maslov DA (2000) Evolution of RNA editing in trypanosome mitochondria. Proc Natl Acad Sci USA **97**: 6986–6993

Stechmann A, Cavalier-Smith T (2003) Phylogenetic analysis of eukaryotes using heat-shock protein Hsp90. J Mol Evol **57**: 408–419

Stevens JR, Noyes HA, Dover GA, Gibson WC (1999) The ancient and divergent origins of the human pathogenic trypanosomes, *Trypanosoma brucei* and *T. cruzi*. Parasitology **118**: 107–116

Strimmer K, von Haeseler A (1996) Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. Mol Biol Evol **13**: 964–969

Vickerman K (1990) Phylum Zoomastigina, Class Kinetoplastida. In Margulis L, Corliss JO, Melkonian M, Chapman DJ (eds) Handbook of Protoctista. Jones and Bartlett, Boston, pp 215–238

Wiemer EAC, Hannaert V, van den Ijssel PRLS, Roy JV, Opperdoes FR, Michels PAM (1995) Molecular analysis of glyceraldehyde-3-phosphate dehydrogenase in *Trypanoplasma borrelli*: an evolutionary scenario of subcellular compartmentation in Kinetoplastida. J Mol Evol **40**: 443–454

Woo PTK, Poynton S (1995) Diplomonadida, Kinetoplastida and Amoebida (Phylum Sarcomastigophora). In Woo PTK (ed) Fish Diseases and Disorders No. 1. CAB International, Wallingford, pp 27–96

Wright A-DG, Li S, Feng S, Martin DS, Lynn DH (1999) Phylogenetic position of the kinetoplastids, *Cryptobia bullocki*, *Cryptobia catostomi*, and *Cryptobia salmositica* and monophyly of the genus *Trypanosoma* inferred from small subunit ribosomal RNA sequences. Mol Biochem Parasitol **99**: 69–76

Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. CABIOS **13**: 555–556